SELF-SYNCHRONOUS INPUT FOR WEARABLES

Thesis
Presented to
The Academic Faculty

By

Shaurye Aggarwal

In Partial Fulfillment
of the Requirements for the Degree
Bachelor of Science in the
College of Computing
Georgia Institute of Technology, Atlanta

Georgia Institute of Technology

May 2019

SELF-SYNCHRONOUS INPUT FOR WEARABLES

Approved by:

Dr. Thad E. Starner, Advisor
School of Interactive Computing
Georgia Institute of Technology

Dr. Gregory D. Abowd
School of Interactive Computing
Georgia Institute of Technology

Date Approved: May 3, 2019

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Wearable devices such as smartwatches and Google Glass commonly employ touch interfaces as input modalities. Performing simple tasks, such as accepting a call or dismissing a notification, using touch screens disrupts the natural flow of conversations and using gestures requires users to get familiar with the gesture vocabulary. This work presents Self-Sync, a gesture interface that enables users to define and perform actions simultaneously in a subtle manner, using a combination of hand, head and leg gestures, without paying attention to the device. Data from tri-axial gyroscopes of an Android smartphone, an Android smartwatch, and a Google Glass is monitored to recognize the intended gesture with high true-positive accuracy (up to 100% for some gestures) and no false positives for higher thresholds. Data was collected through an in-lab user study and gestures were analyzed on the basis of accuracy, taskload, user preference, social acceptability, and user feedback. After picking two most ideal gestures based on the evaluations, we designed an in-the-wild experiment to further test our interface. Applications and current limitations of the system are further discussed.

CHAPTER 1

INTRODUCTION

Wearable devices, such as smartwatches and Google Glass, are becoming increasingly popular among users in part due to their portability and readily available access to content. However, as wearable devices stay on the user compared to traditional computers, input to such devices through touch are noticeable and tend to disturb social interactions, such as a conversation with others. This limits the acceptability of popular voice based input methods and makes designing appropriate interactions for wearables more complicated. For users carrying multiple devices, using voice based input might trigger response in multi-device scenario e.g. when a user operates a smartwatch and a smartphone. In some cases, where user's hands are occupied, touch based and hardware button based interactions become more challenging to perform for basic micro-interactions such as accepting/rejecting a call, dismissing a notification, activating the screen or initializing input.

Previously, various input wearable technologies have been designed to address this issue for input interactions by recognizing a set of gestures, using various sensing modalities worn on fingers [1, 2], wrists [3, 4, 5] and arm [6, 7]. While some of the approaches make use of the sensors integrated with the wearables (e.g., accelerometer, magnetometer, and gyroscope) [8, 9, 10] others require complicated hardware setup to support a diverse range of gestures, which makes the system cumbersome to wear [6, 7, 4, 5] or they require extra wearables [1, 2], driving users to carry more devices. Few gestures are limited by their memorability and learning curve [11, 12, 13, 14]. Instead of directly recognizing gestures as aforementioned systems did, few projects focused more on localizing the position of a finger or arm for continuous input [15, 16, 17]. While continuous input systems do address the gesture memorability problem to a certain extent, they require external sensing capabilities or are not suitable for micro-interactions such as notification dismissal or input initialization.

In contrast to directly recognizing multiple gestures from a vocabulary or tracking continuous motion of a body part, few approaches focused on synchronous interfaces, where a user expresses intent by performing a motion that is synchronized with the target stimulus and hence the gestures require minimal familiarization and memorization. Such systems have been developed for devices such as smartwatches, using additional hardware to track eye-gaze [18], finger [19] and hand-motion [20]. For synchronous interfaces, generally different selection options on the user interface are mapped to suitable gestures. In most cases, user is expected to perform a motion that is synchronized with a path (target stimulus) presented through the UI. While this kind of mapping attenuates memorization needs, the user still needs to pay attention to the device. Wu et al. found users were able to provide one-handed smartwatch only input, using only haptic stimuli and proposed an eyes-free synchronous gestures by using two body parts simultaneously in the absence of a stimulus [21]. The question remains whether the user can provide input to a wearable, such as a smartwatch or a Google Glass, by performing a motion that is synchronized with another user-defined motion.

Self-Sync is a subtle, gaze-free self-synchronous body based interface that used a combination of hand, head, and leg motion. Self-Sync enables the user to define the stimulus and simultaneously perform the gesture and thus is expected to require lower concentration. Moreover, due to the uncommon coordinated movement between two body parts, we hypothesize that a self-synchronized interaction interface would allow recognition with higher accuracy and lower false-positives. It exploits triple axis sensors integrated in off-the-shelf smartwatches, head worn displays, and smartphones, which allows accurate and coordinated gesture sensing in a multi-device scenario.

This work first presents Self-Sync's gesture design and implementation of the Self-Sync gesture interaction and the underlying recognition method. We explored SelfSync gestures in an offline in-lab user study, conducted in the United States. Results from the study are presented and discussed. This work then briefly describes an online in-the-wild experiment, being conducted in South Korea, to test two optimal SelfSync gestures that were chosen based on our evaluations. Finally, current limitations and future research avenues are presented and discussed.

CHAPTER 2

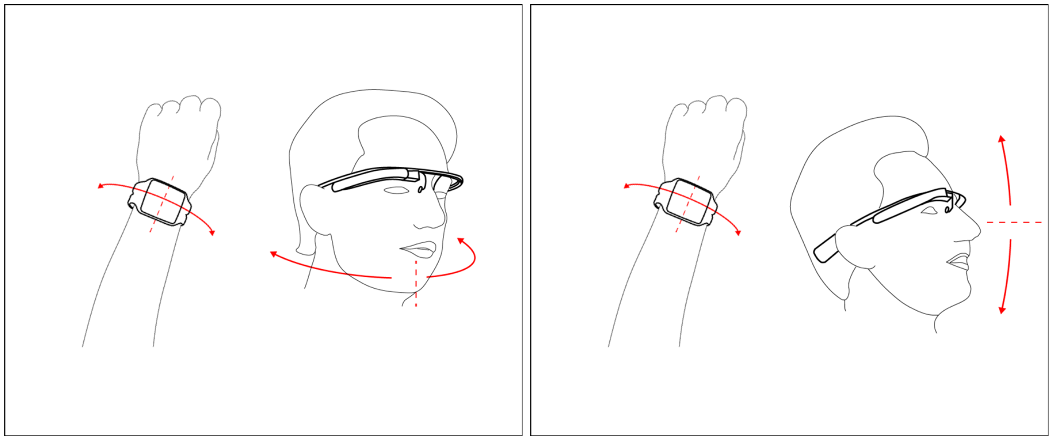
INTERACTION DESIGN AND SYSTEM OVERVIEW

2.1 Interaction Design

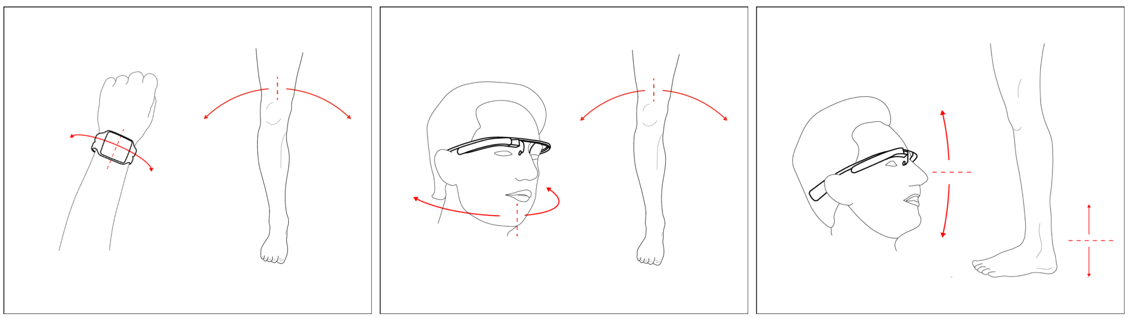
SelfSync gestures are meant to be recognized by exploiting synchronous motion of multiple body parts. To choose gestures for our initial evaluation, we looked at common locations, where a user might keep or wear a sensor-enabled device.

> Leg - Smartphones are often placed in pockets when idle, where "they are well positioned for capturing information about the orientation and movement of the leg" and are capable of receiving IMU input for activity recognition and fitness tracking. Moreover, leg gestures have previously been recognized using smartphones kept in users' pockets [22].
>
> Hand - Many approaches in the past have explored devices that are worn on the finger, wrist, hand, or arm. Devices such as smartwatches, smart rings, and sensor-enabled armbands are adept at capturing gross and subtle, arm, finger, and wrist movements. Most off-the-shelf Android smartwatches come with integrated 9-axis IMU sensors and are capable of processing and transmitting sensor data.
>
> Head - Various devices that can be worn on the head, such as headphones, ear-buds, smart eye-wear, and head worn displays like Google Glass, include sensors such as accelerometers, gyroscopes, and electrooculography (EOG) electrodes which can be used to characterize head pose, facial expressions and eye movement.

We chose five gestures that can be performed using the aforementioned body parts: leg left-right (toe rotation) and up-down (dorsiflexion); head left-right and up-down; and hand twist. For Self-Sync, we initially considered 7 combinations of these gestures: (a) *hand*



(a) Hand Twist & Head Left-Right (b) Hand Twist & Head Up-Down



(c) Hand Twist & Leg Left-Right (d) Head Left-Right & Leg Left-Right (e) Head Up-Down & Leg Up-Down

Figure 2.1: Self-Sync Gestures

twist & head left-right, (b) *hand twist & head up-down*, (c) *hand twist & leg left-right*, (d) *head left-right & leg left-right*, (e) *head up-down & leg up-down*, (f) *hand twist & leg up-down*, and (g) *head left-right & leg up-down*.

Prior to our in-lab user study, we ran a similar pilot study on two subjects from our lab. The last two gestures (*hand twist & leg up-down* and *head left-right & leg up-down*) had lowest true-positive accuracy and the subjects found them to be much harder than the rest of the gestures. Hence, we omitted them from the gesture set (Figure 2.1) for our evaluation and decided to explore the first five gestures in our in-lab evaluations.

2.2 System Overview

2.2.1 Hardware

Most commodity smart devices come with Inertial Measurement Unit(IMU) sensors to track devices, motion and orientation. We use a Sony Smartwatch 3 SWR50 that runs on a Quad ARM A7 1.2 GHz processor, Google Glass XE that runs on a 1.2Ghz Dual(ARMv7) processor, and an Android smartphone to capture wrist, head, and leg movement respectively. Each device possesses a 9-axis IMU.

2.2.2 Data Processing

For capturing motion information from all body parts to classify SelfSync gestures, we used the in-built gyroscopes. We sampled the watch gyroscope at 33Hz and the glass and phone gyroscopes at 100Hz. The data was sent to a central server for processing through UDP via WiFi.

For the offline data gathering, we used a Macbook Pro and for the in-the-wild study, we used the phone as the central device. We implemented the offline system in Python with Pygame for visualization and Scikit-learn for training machine learning classifiers and for the online system we used the Android SDK and Weka.

We exploited the correlation coefficient to develop a simple threshold-based system, which was used during data collection, as a first step towards real-time recognition. For calculating the correlation, we first segmented the data from each device into 1.5-second windows and ran a Principal Component Analysis that converted three axes of each gyroscope to one dominant axis, highlighting the synchronous gesture. Then, we ran cross-correlation to match time series from two different devices and ultimately calculated the Pearson correlation coefficient. We did this process for each pair: head & hand, head & leg, and hand & leg. With this system, we were able to distinguish body parts, but not the exact gesture.

To classify different SelfSync gestures, we created our own feature vector to train a Random Decision Forest classifier. We extracted 17 total features, from each device, which includes 11 statistical features - (i) maximum, (ii) minimum, (iii) difference of maximum and minimum, (iv) mean, (v) standard deviation of the raw signal, (vi) root mean square value of the segmented signal, and (vii) maximum, (viii) minimum, (ix) mean, (x) standard deviation, and (xi) root mean square value of the discrete difference between consecutive segmented signal - and six other features - number of (xii) positive and (xiii) negative peak in the signal, the absolute value of PCA components for (xiv) x, (xv) y, (xvi) z-axis, and (xvii) the biggest axis among three axes. We extracted seven more features from the different combinations of all body parts including each pair's raw correlation value, each pair's correlation value compensated by cross-correlation, and the pair which has the maximum compensated correlation value. Our feature vector included $3 \times 17 + 7 = 58$ features for each segmented window.

CHAPTER 3

TRUE POSITIVE DATA COLLECTION AND INITIAL GESTURE EVALUATION

To collect true positive data for the classifier and discern the differences between the five chosen gestures, we conducted a within-subject offline study in our laboratory. In this section, we describe the experimental procedure and discuss our evaluations of Self-Sync gestures based on user response.

Upon arrival, each participant was asked to mount the Google Glass, wear the smartwatch on their left wrist, and keep the smartphone in their right-leg pocket Each device was already connected to the laptop and ready to publish sensor data to the central server. The participants wore the devices while reading and signing the consent form and sensor data was transmitted to the server during this phase. Participants then performed the gestures across two activities (sitting and standing) and answered a questionnaire in the end to further evaluate difficulty, preference, and acceptability of the chosen gestures. Sensor data from all the three wearables was saved as log files as false-positive and true-positive data respectively. The study lasted approximately one hour per participant.

3.1 Participants

We conducted the study with 10 students (all male, ages 20-25) from our institution in the United States, recruited via word of mouth. Two participants regularly used wearables (smartwatches) for tasks such as notification updates, picking calls, controlling music player and monitoring health. Participants received $10 compensation for their time.

3.2 Procedure

3.2.1 False-Positive Data Collection

Once comfortable with the three devices, the participants walked around the lab, while reading the consent form. The experimenter walked alongside the participants, while explaining the purpose and procedure of the study. Approximately, five minutes of false-positive data was collected from each participant, before they were ready to be trained on the gestures. In totality, we collected 45 minutes of false-positive data.

3.2.2 True-Positive Data Collection

For training, the experimenter demonstrated each gesture in no specific order. The participants were asked to practice till they could confidently perform the gesture. A visual feedback of gesture detection by the threshold-based classifier - which was also used in the study - was shown during this phase. The participants then got acquainted with the experimental setup.

For the main study, each gesture was presented in random order and participants were asked to perform the target gesture as naturally as possible. Through pilot studies, we found that gestures, especially the ones involving the leg, had varying levels of difficulty depending on the user posture. Hence, we tested the participants over two conditions - standing and sitting. Participants were randomly chosen to either sit or stand during the first set of trials.

Users began performing the gesture during a five second “warmup” period that commenced after acknowledging the target gesture by pressing a key on the keyboard. Followed by the “warmup” period was a two second “recording” period that began automatically after the “warmup” period ended. The participants were notified of the start and end of the “recording” period by a beep. Here, the participants were asked to perform the gesture as accurately as possible. Another beep marked the end of the two second period.

During pilot studies, we also found that participants can get confused between the gestures, especially during the later stages of the experiment and make unintentional errors such as moving all the three body parts together. To later remove gestures that were performed incorrectly from the true positive data, the experimenter took note of the trials where participants performed the incorrect gesture.

The accelerometer and magnetometer data from the three devices along with their respective timestamps were stored during both set of trials independently. Sensor data collection from all three devices started when each set of trials was initiated and ended when the participant finished performing the last gesture for each set The order and timestamps for the target gestures were also recorded for each set and later used to match sensor data with their respective gesture labels. Participants performed five repetitions for five gesture across two conditions giving a total of 50 gestures per participant.

In the case of synchronous gesture, a user generally expresses intent through performing a motion that is synchronized with the target stimulus. For Self-Sync, the user defines a stimulus and expresses intent simultaneously by synchronously moving a combination of their body parts. Hence, for each SelfSync gestures we expected participants to prefer a main body part (faster motion) which they follow using their other body part (slower motion). The true-positive data collected from the study was used to find body part preference in each gesture. Furthermore, to assess how well the body parts are synchronized with each other among different gestures we compared the correlation values.

3.2.3 Questionnaire

After data collection was complete, the participants completed a questionnaire that included, (i) for every gesture, (a) a social acceptability questionnaire [23] and (b) a NASA TLX along with (ii) general demographic questions, (iii) qualitative questions such as likes and dislikes about SelfSync gesture set and (iv) a gesture preference questionnaire. The gesture preference questionnaire asked the participants to rank the gestures in the order of their preference and was used to find which gestures are more preferred by the users. The social acceptability questionnaire aims to evaluate social acceptability of each gesture with respect to the audience they perform the gesture in front of and location they perform the gesture in and was used to understand social acceptability of each gesture. The NASA TLX was used to assess the overall difficulty of each gesture with respect to others and was administered using the official NASA TLX app. Data from the questionnaire was used to get further recommendations on choosing the final gesture set

3.3 Result

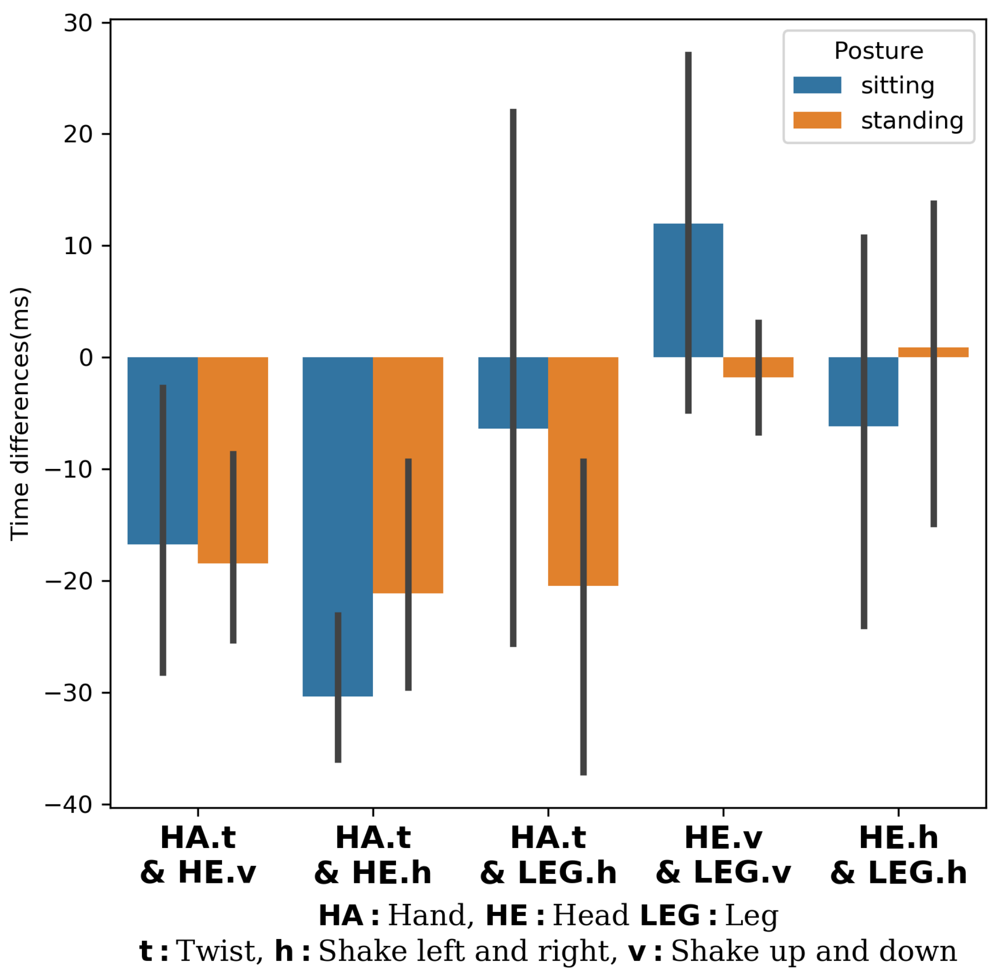


Figure 3.1: Average difference between different body parts in each frame

3.3.1 Body-Part Preference

A paired t-test on average time difference between different body parts for all 1-second frames of each gesture with Benjamini-Hochberg correction revealed hand was faster than head for *hand twist & head left-right* and *hand twist & head up-down* ($p < 0.0001$). Hand is more likely to be faster for *hand twist & leg left-right*. For leg and head gestures, it's more likely that head is faster while sitting and standing, except for *head left-right & leg up-down* where leg is more likely to be faster while standing. The results are shown in Figure 3.1.

3.3.2 Correlation Comparison

Using a pairwise t-test with Benjamini-Hochberg correction, we found a significant difference in correlation values for most gesture pairs but three: among *hand twist & head left-right*, *hand twist & head up-down*, and *head left-right & leg left-right*. Correlation for *hand twist & leg left-right* and *head up-down & leg up-down* were significantly lower than the three aforementioned gestures ($p < 0.01$) and among the two, the latter had a higher correlation value ($p < 0.01$) (see Figure 3.2). Correlation values were significantly higher in cases where participants were standing (p=0.01) (see Figure 3.3).

3.3.3 Taskload

Using one-tailed paired student t-test with Benjamini-Hochberg correction ($\alpha = 0.05$), we found the summed raw scores across the six dimensions for *hand twist & head left-right*, to be significantly less than *hand twist & head up-down* ($t(9) = -3.898$, $p < 0.05$, $Cohens_d = 1.233$) (Table 3.1). Hand Twist and & Head Left-Right had the lowest mean task load among all the gestures and Head Up-Down and Leg Up-Down had the maximum. Among leg gestures, Hand Twist and Leg Left-Right had lower mean overall score than Head & Leg Up-Down and Head & Leg Left-Right (Table 3.2). Results are shown shown in Figure 3.4 and Figure 3.5.

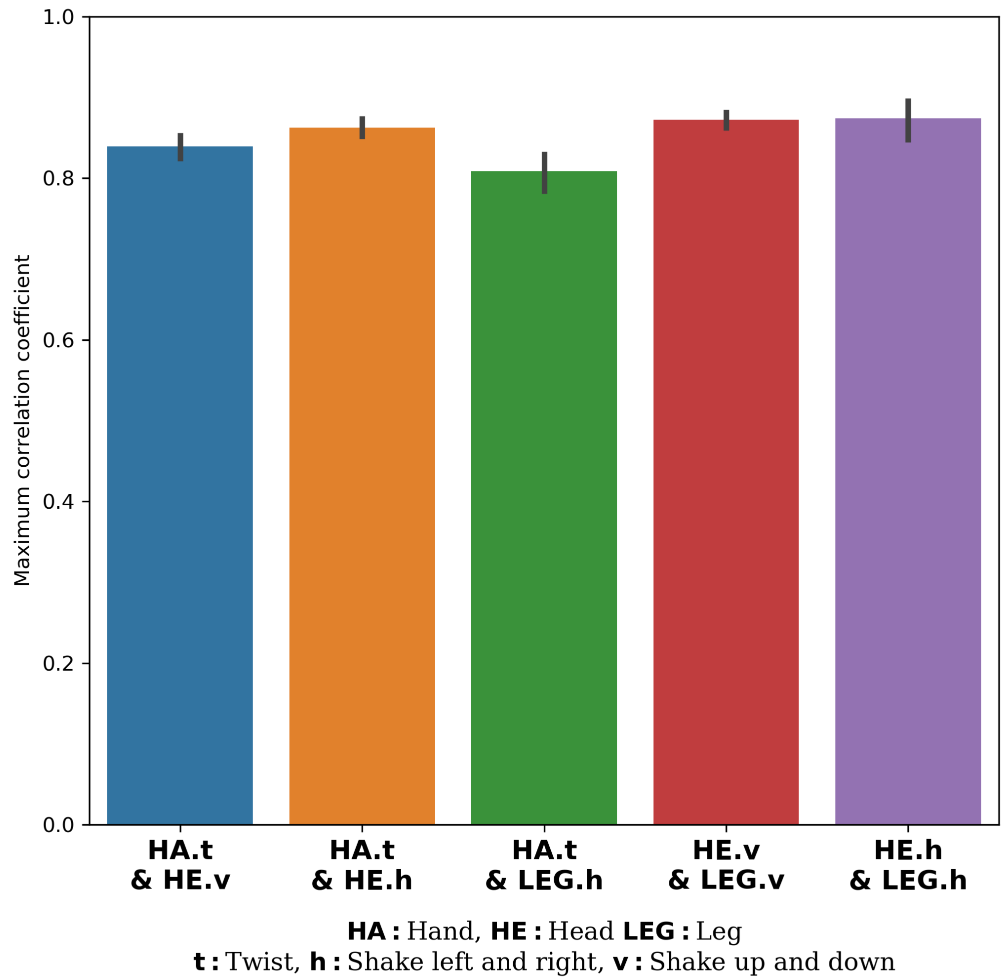


Figure 3.2: Correlation of each gesture

	Hand Twist & Head Up-Down	Hand Twist & Head Left-Right	Hand Twist & Leg Left-Right	Leg Left-Right & Head Left-Right	Leg Up-Down & Head Up-Down
Hand Twist & Head Up-Down	-	(3.898, 0.036)	(0.661, 0.656)	(0.021, 0.984)	(-0.483, 0.712)
Hand Twist & Head Left-Right	-	-	(-0.852, 0.656)	(-1.845, 0.327)	(-2.494, 0.171)
Hand Twist & Leg Left-Right	-	-	-	(-1.217, 0.509)	(-1.374, 0.507)
Leg Left-Right & Head Left-Right	-	-	-	-	(-0.693, 0.656)

Table 3.1: t-statistics and p values obtained by paired Student t-test with Benjamini-Hochberg correction ($\alpha = 0.05$)

3.3.4 Gesture Preference

While, using pairwise one-tailed Wilcoxon Signed Rank test with Benjamini-Hochberg correction ($\alpha = 0.05$) we found no significant difference between preference ranking for

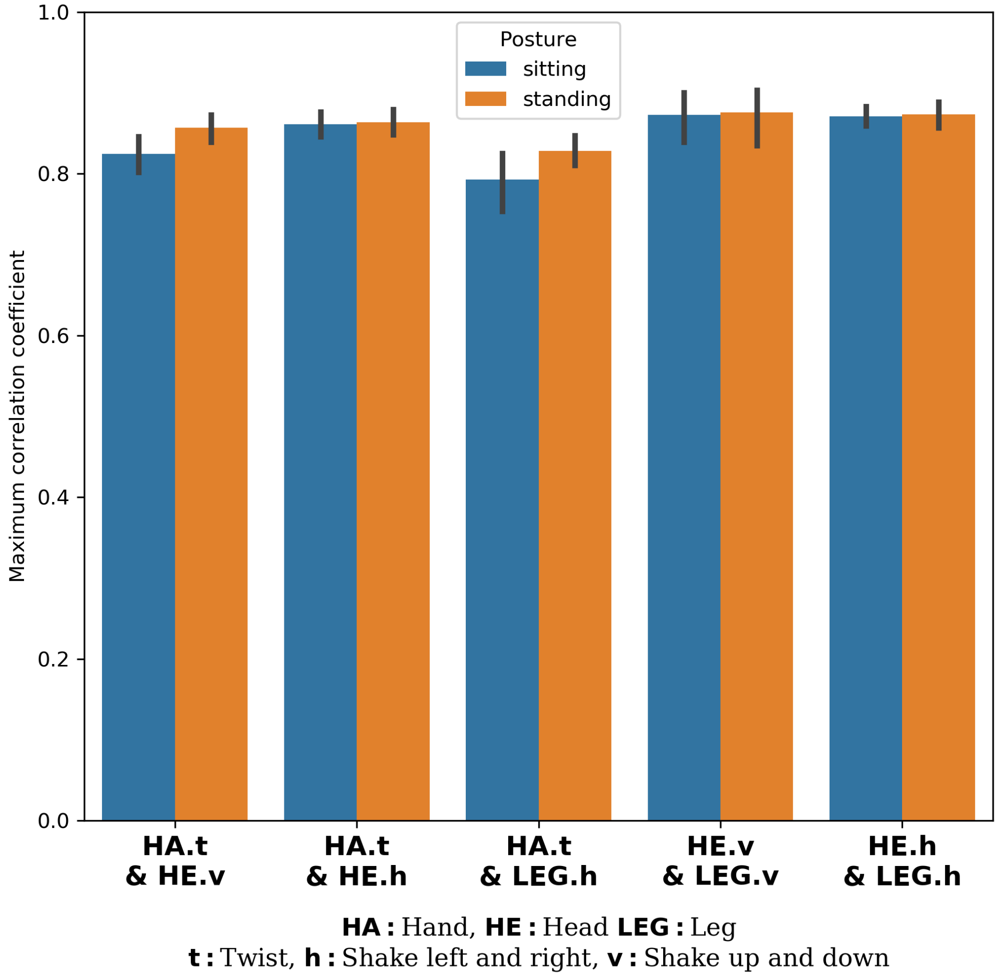


Figure 3.3: Correlation of each gesture in different posture

	Mean	Standard Deviation
Hand Twist & Head Up-Down	244.5	103.6
Hand Twist & Head Left-Right	171.0	84.5
Hand Twist & Leg Left-Right	208.0	120.2
Leg Left-Right & Head Left-Right	243.5	104.3
Leg Up-Down & Head Up-Down	266.0	117.6

Table 3.2: Statistics corresponding to the Overall Taskload for all Gestures

gestures, Leg Up-Down & Head Up-Down was the only gesture not rated as "most preferred" by any participant. Rankings for *hand twist & head up-down* ($std_a = 2$) were more divided than other gestures ($std_b = 1.26$, $std_c = 1.14$, $std_d = 1.49$, and $std_e = 1.07$); the participant either preferred it the least ($n_1 = 4$) or preferred it highly ($n_4 = 1$, $n_5 = 5$).

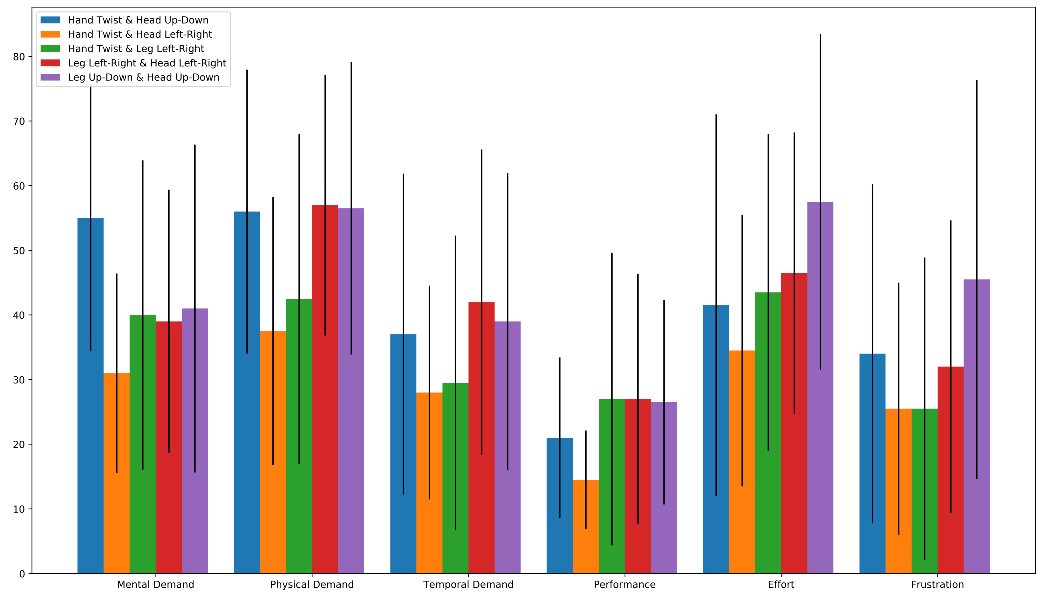


Figure 3.4: Means for each scale between all gestures

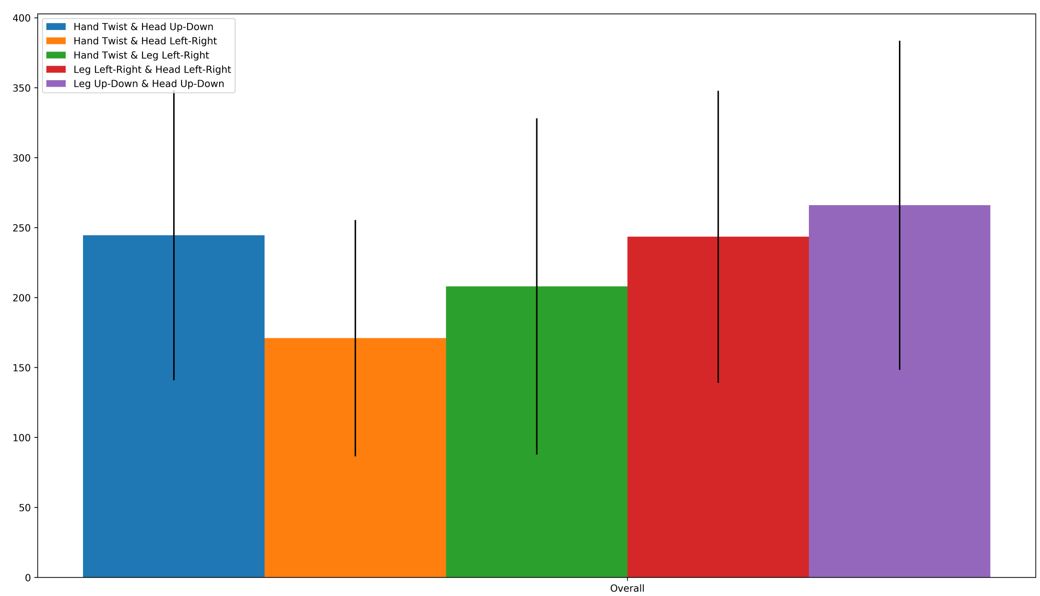


Figure 3.5: Mean overall taskload for all gestures

Gestures involving leg had lower mean rankings ($\mu_c = 2.8$, $\mu_d = 2.7$, and $\mu_e = 2.6$) compared to those involving only hand and head ($\mu_a = 3.3$ and $\mu_b = 3.6$).

3.3.5 Social Acceptance

We summed the number of contexts participants agreed to perform the gesture in for each location and audience category to get an acceptability score for all gestures.

For location, gestures requiring both head and leg had lower mean acceptability compared to gestures requiring hand twist. *Hand twist & leg left-right* had the highest acceptability among the gestures involving legs. Between *hand twist & head up-down* and *hand twist & head left-right*, more participants were likely to perform the latter at various locations. All participants were willing to perform the gestures at *home* and were least likely to perform the gestures when they were *driving*.

In terms of audience, though we did not find any significant difference between gestures' acceptabilities, the average acceptability for leg-based gestures (< 4) was less than those not involving leg (> 4). On a closer look, all participants were willing to perform the gestures when alone and were more likely to perform gestures in front of *partner* and *family* compared to *strangers* and *colleagues*.

3.3.6 Qualitative Feedback

When asked to describe their experience with the system, one participant reported that moving the head up and down was the "most tedious" among all body parts and they didn't like their "vision to be dictated by head gestures." Although few participants reported that it didn't feel natural to move the leg in many occasions, especially while standing, one felt leg is "easier to move" compared to head. One of them further mentioned that twisting wrist "felt the most natural" as it was "very easy to complete without raising arms too much." They believed that while Head and Leg gestures felt "rhythmic," they struggled with Leg and Head gesture combos occasionally and called them "more uncomfortable." Two participants specifically mentioned that gestures were easy to learn.

3.4 Discussion

From initial pilot studies, we learned that gestures involving similar motion were easier to perform and synchronize compared to the rest. This knowledge became the prime reason to exclude gestures - *head left-right & leg up-down* and *hand twist & leg up-down*. We saw a similar trend in our results: among *hand twist & head left-right* (similar motion) and *hand twist & head up-down*, the former was easier as shown by NASA TLX and while not significant, was more socially acceptable in terms of location and audience and more preferred by a majority of our subjects, indicating that it's a more suitable choice for an initialization gesture among hand and head gestures.

Gestures with only legs or head such as foot tapping have been said to be more uncomfortable compared to other gestures involving hand [23]. In the qualitative feedback, participants felt that leg or head gestures were more challenging, while the hand twists were relatively less demanding. Evaluation of the mean time difference per frame for SelfSync gestures involving hand, revealed that the participants tried to follow their hand with their head and legs. We believe this might be due to participants' preference for hand gestures over other body parts. Among head and leg gestures, participants generally preferred leg except in the case when they are standing and executing Leg Up-Down. This might be due to the fact that Head Up-Down is much simpler compared to the harder dorsiflexion motion when legs are supporting the weight of the body. Despite hand gestures being more comfortable, results indicate Hand Twist and Leg Left-Right had poor synchronization compared to all gestures, especially both the head and leg gestures. Participants reported to find rhythm in Head and Leg gestures which might have led to better synchronization of the two body parts. However, *hand twist & leg left-right*, while not significant, had the second lowest mean taskload and a higher mean location acceptance compared to other leg gestures.

In pilot studies, subjects also found Leg Left-Right to be more comfortable than Leg

Up-Down, supporting previous findings by Scott et al. on foot gestures [22] (toe rotation vs. dorsiflexion respectively). Our results show that *head up-down & leg up-down* had significantly lower correlation compared to both Head and Hand gestures and *head left-right & leg left-right* implying the synchronization between body parts was worse for Leg Up-Down gestures. None of the participants picked it as ”most preferred” gesture further indicating them to be less comfortable, as mentioned by a few participants.

These results gave us a better idea for choosing gestures for our in-the-wild study. Hand and Head gestures, generally, were more correlated, comfortable, preferred and socially acceptable than gestures involving leg. This fact led us to believe that *hand twist & head left-right* might be the strongest candidate among all SelfSync gestures. Among gestures involving legs, while head and leg gestures had higher correlation values, we believe *hand twist & leg left-right* would be more ideal as an initialization gesture due to lower mean taskload, higher mean location social acceptance, and greater comfort.

CHAPTER 4

SELFSYNC CLASSIFIER

We removed eight outliers from our data set and created a new dataset by taking one second windows from the raw true positive data and randomly choosing the same number of one second windows from the raw false positive data collected from our in-lab study. Next, we trained a Random Decision Forest classifier on this new dataset and evaluated the gestures based on accuracies for user-independent, user-dependent, and user adaptive models.

4.1 Offline Training Result

4.1.1 User-Independent Accuracy

We built user-independent models by training the classifier on data from a participant and testing it across data from all the other participants. In our results (Figure 4.1), *hand twist & leg left-right* had the highest average accuracy of 96.8%, followed by *hand twist & head up-down* with 96.8% accuracy. *Head up-down & leg up-down* was recognized with 88.2% average accuracy. *Hand twist & head left-right* and *head left-right & leg left-right* had the lowest average accuracies of 82.2%.

4.1.2 User-Dependent Accuracy

User-dependent models were trained using a subset of each user's data and then was used to recognize their own gestures. We performed 3-fold cross validation on each user's data and achieved 98.5% average accuracy for *hand twist & leg left-right*. Average accuracies for *hand twist & head up-down* and *head up-down & leg up-down* reached 95.5%. *Head left-right & leg left-right* and *hand twist & head left-right* were recognized with average accuracies of 95% and 92.5%. Results are shown in Figure 4.2.

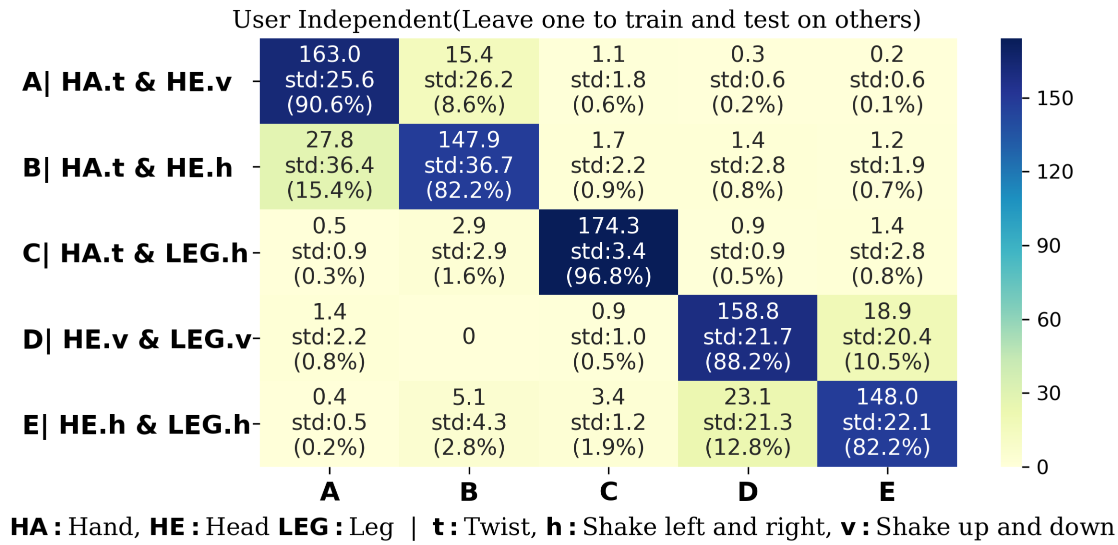


Figure 4.1: User Independent accuracy with all study data(leave-one-out)

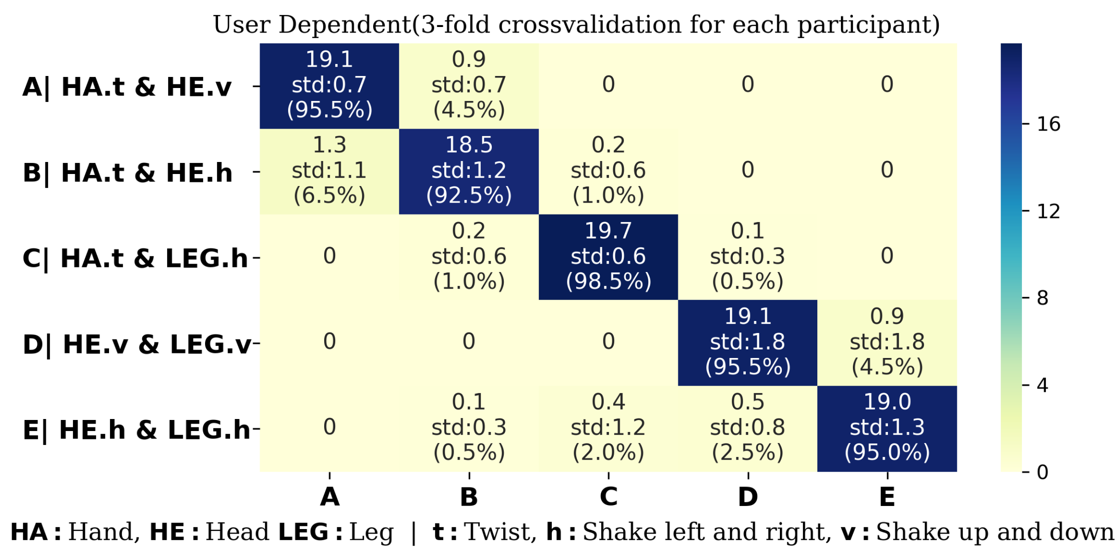


Figure 4.2: User Dependent accuracy with all study data(cv=3)

4.1.3 User-Adaptive Accuracy

Training a user-independent model with additional training instances from a specific user results in a user-adaptive model. By performing 10-fold cross validation on the whole dataset, we achieved 100% average accuracies for *head & leg up-down* and *hand twist &*

head up-down. *Hand twist & head left-right* reached average accuracy of 98.5% followed by *head & leg left-right* (98.5%) and *hand twist & leg left-right* (95%). Results are shown in Figure 4.3.

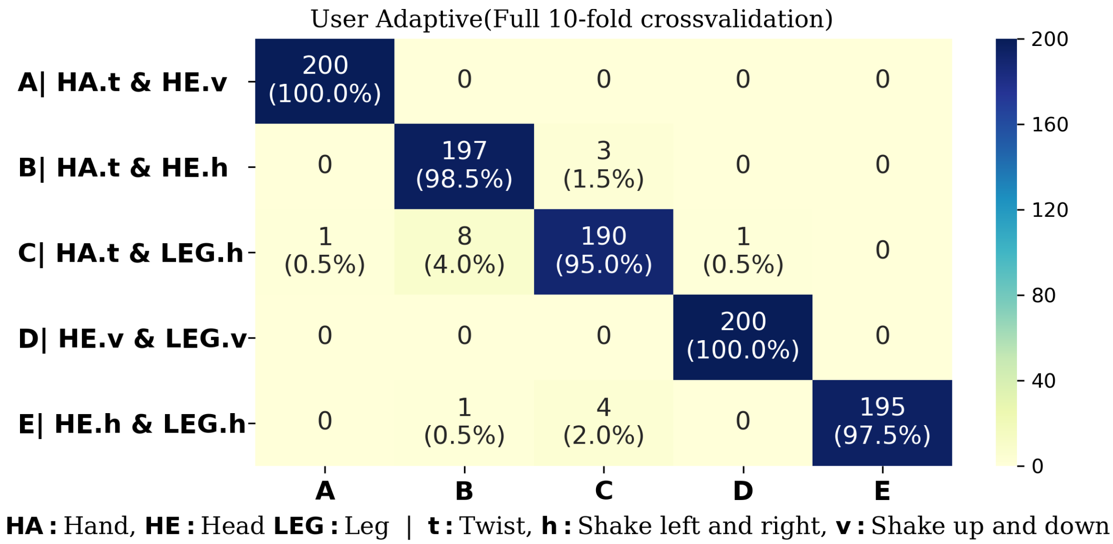


Figure 4.3: User Adaptive accuracy with all study data(cv=10)

4.1.4 False-Positives

We tested the user-adaptive models for false positives at four confidence level thresholds: 0.6, 0.7, 0.8, and 0.9. It detected false-positives with an accuracy of 97.5% without considerable change in the accuracies. At 0.6 confidence threshold, when we ran all the false-positive data we collected at the beginning of our user study, false-positive rate was 5.29 errors per hour. Our classifier detected four false-positives, as either *hand twist & head left-right* or *head up-down & leg up-down*. For all other thresholds, SelfSync classifier was robust to false-positives, having an error rate of zero per hour.

4.2 Discussion

A gesture for input initialization should achieve high accuracy with lower number of false positives to be practical for everyday use. SelfSync gestures are recognized with high accuracy rates in user-independent ($min_{acc} = 82.2\%$, $max_{acc} = 96.8\%$), user-dependent ($min_{acc} = 92.5\%$, $max_{acc} = 98.5\%$) and user-adaptive cases ($min_{acc} = 95\%$, $max_{acc} = 100\%$). User-independent models since are trained without any training instances from the testing user, had lower accuracies compared to user-dependent and user-adaptive models. User-dependent models were mainly confused between gestures using the same body part combinations (hand & head gestures and head & leg gestures) and this uncertainity is lost in user-adaptive models by training on more gesture data from other users, leading to higher accuracies in latter's case. However, our user-adaptive models recognized a few false-positive instances as gestures involving leg. Classification of non-gesture data as leg gestures could be due the fact that users were mainly involved with walking and reading the consent form during the false-positive data collection phase of the user study. Moreover, for thresholds ≥ 0.7 , our classifier was capable of differentiating every day action from actual gestures and resulted in zero false-positives without affecting the actual accuracies considerably. We believe that everyday actions rarely include synchronous motion of two body parts and hence SelfSync is robust to false-positive.

Among hand and head gestures, *hand twist & head up-down* has higher accuracy in user-independent and user-dependent models. However, since *hand twist & head up-down* achieved similar and higher accuracies in user-dependent and user-adaptive cases respectively and was more suitable in terms of taskload, acceptability, and user preference, we think it is a better gesture among the two.

While between head and leg gestures, *leg up-down* had worse synchronization compared to *leg left-right*, it achieved better accuracy in all cases. Moreover, *hand twist & leg left-right* had the lowest correlation value but achieved the best accuracy among the leg gestures in all cases. Hence, when results from the previous section are taken into account, *hand twist & leg left-right* comes out as the most optimal gesture among SelfSync gestures involving leg.

CHAPTER 5
IN-THE-WILD EXPERIMENT

In the previous study, we explore the possibility of SelfSync as a new interaction method. The results highlighted the differences between the gestures and indicated that Self-Sync is a practical gaze-free, multi-device interaction interface. To evaluate SelfSync in real world settings, we designed an in-the-wild study for SelfSync with the two optimal self-synchronous gestures we found in the previous study: *hand twist & head left-right* and *hand twist & leg left-right*. Participants are asked to perform one of the two gestures when prompted through a notification randomly. For notifications, we play an audio and use haptic feedback for 20 seconds or till the classifier recognizes the target gesture correctly. We are running our Random Decision Forest classifier on the phone, during the whole experiment to collect both, the false-positives as well as the true-positives. Before the start of the experiment, each participant took 5-10 minutes to get comfortable with the gestures and the setup. For the next one hour, the participants wore all three devices and carried out their usual work, while performing the gesture when prompted. Finally, the participants answer questions pertaining to social acceptability, gesture taskload and gesture preference. We are collaborating with researchers at Korea Advanced Institute of Science and Technology (KAIST), South Korea, on SelfSync. Currently, the in-the-wild experiment is being conducted in South Korea where participants are also paid approximately $10 for their time. The preliminary results look promising as both chosen gestures are being recognized with high accuracy without triggering any false positives. We intend to publish the final results in a full-length paper.

CHAPTER 6

LIMITATIONS AND FUTURE WORK

SelfSync is a gesture interface that enables users to perform gaze-free interactions with their phones comfortably using a combination of hand, head, and leg motion. Our classifier, though is able to recognize SelfSync gestures with high accuracy, the accuracies for leg gestures are lower due to the small motion of the leg compared to other body parts. Currently, we only used data from the devices' gyrscopes to classify the gestures as including accelerometers resulted in a drop in true-positive accuracy. However, one can also only exploit the phone's accelerometer to improve the detection of the gestures involving leg.

Another limitation of SelfSync is that it currently only supports right-leg pocket, but in the real world user tend to switch pockets often. Some might even get confused between the pockets when trying to locate the phone. This confusion is problematic as the users then might perform leg gestures using the wrong leg. This confusion is also problematic in cases where the front pockets are too small to accommodate the smartphones which again might hinder the capturing of leg motion. Additional work can be done to add support for recognizing leg gestures from more locations.

Lastly, our SelfSync recognizer is able to classify gestures when performed with motions of less magnitude. Since, the SelfSync gestures themselves can be performed subtly, we predict that SelfSync supports subtle interactions. There are many benefits to having subtle interactions. It protects users privacy, tends to be more acceptable, and allows users to perform the gesture in much more confined physical space [24]. To quantify SelfSync gestures subtleness, we intend to conduct a noticeability study.

CHAPTER 7
CONCLUSION

In this work, we presented a multi-device self-synchronous interaction interface that allows gaze-free input requiring lower concentration. In contrast to many common gesture interfaces, the self-synchronous interface allows the user to define a stimulus and perform the action simultaneously and hence, is independent of a vocabulary, making it easier to remember. Moreover, we found the gesture to be recognized with high true-positive accuracy ($> 90\%$ for user adaptive and user-dependent models). SelfSync's false-positive rates were low due to the uncommon motion of the gesture. SelfSync, hence, is a suitable choice for micro-interactions such as declining a call, dismissing a notification, or activating device for a different gesture input. We evaluated the gestures through an in-lab true-positive study and found *hand twist & head left-right* and *hand twist & leg left-right* to be the most viable gestures based on recognition accuracy, social acceptabilty, and user preference. Viability of the gestures were further discussed based on user feedback. Finally, we briefly described the ongoing in-the-wild experiment that aims to examine performance of the two aforementioned gestures in real life.

REFERENCES

[1] C. Zhang, X. Wang, A. Waghmare, S. Jain, T. Ploetz, O. T. Inan, T. E. Starner, and G. D. Abowd, "Fingorbits: Interaction with wearables using synchronized thumb movements," in *Proceedings*, ser. ISWC, Maui, Hawaii: ACM, 2017, pp. 62–65, ISBN: 978-1-4503-5188-1.

[2] C. Zhang, A. Waghmare, P. Kundra, Y. Pu, S. Gilliland, T. Ploetz, T. E. Starner, O. T. Inan, and G. D. Abowd, "Fingersound: Recognizing unistroke thumb gestures using a ring," *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, vol. 1, no. 3, 120:1–120:19, Sep. 2017.

[3] A. Dementyev and J. A. Paradiso, "Wristflex: Low-power gesture input with wrist-worn pressure sensors," in *Proceedings*, ser. UIST, Honolulu, Hawaii, USA: ACM, 2014, pp. 161–166, ISBN: 978-1-4503-3069-5.

[4] H. Xia, T. Grossman, and G. Fitzmaurice, "Nanostylus: Enhancing input on ultra-small displays with a finger-mounted stylus," in *Proceedings*, ser. UIST, Charlotte, NC, USA: ACM, 2015, pp. 447–456.

[5] R. Xiao, G. Laput, and C. Harrison, "Expanding the input expressivity of smart-watches with mechanical pan, twist, tilt and click," in *Proceedings*, ser. CHI, Toronto, Ontario, Canada: ACM, 2014, pp. 193–196, ISBN: 978-1-4503-2473-1.

[6] C. Harrison, D. Tan, and D. Morris, "Skinput: Appropriating the body as an input surface," in *Proceedings*, ser. CHI, Atlanta, Georgia, USA: ACM, 2010, pp. 453–462, ISBN: 978-1-60558-929-9.

[7] T. S. Saponas, D. S. Tan, D. Morris, R. Balakrishnan, J. Turner, and J. A. Landay, "Enabling always-available input with muscle-computer interfaces," in *Proceedings*, ser. UIST, Victoria, BC, Canada: ACM, 2009, pp. 167–176, ISBN: 978-1-60558-745-5.

[8] G. Laput, R. Xiao, and C. Harrison, "Viband: High-fidelity bio-acoustic sensing using commodity smartwatch accelerometers," in *Proceedings*, ser. UIST, Tokyo, Japan: ACM, 2016, pp. 321–333, ISBN: 978-1-4503-4189-9.

[9] G. Reyes, D. Zhang, S. Ghosh, P. Shah, J. Wu, A. Parnami, B. Bercik, T. Starner, G. D. Abowd, and W. K. Edwards, "Whoosh: Non-voice acoustics for low-cost, hands-free, and rapid input on smartwatches," in *Proceedings*, ser. ISWC, Heidelberg, Germany: ACM, 2016, pp. 120–127, ISBN: 978-1-4503-4460-9.

[10] C. Zhang, A. Bedri, G. Reyes, B. Bercik, O. T. Inan, T. E. Starner, and G. D. Abowd, "Tapskin: Recognizing on-skin input for smartwatches," in *Proceedings*, ser. ISS, Niagara Falls, Ontario, Canada: ACM, 2016, pp. 13–22, ISBN: 978-1-4503-4248-3.

[11] D. Ashbrook, P. Baudisch, and S. White, "Nenya: Subtle and eyes-free mobile input with a magnetically-tracked finger ring," in *Proceedings*, ser. CHI, Vancouver, BC, Canada: ACM, 2011, pp. 2043–2046, ISBN: 978-1-4503-0228-9.

[12] C. Harrison and S. E. Hudson, "Abracadabra: Wireless, high-precision, and unpowered finger input for very small mobile devices," in *Proceedings*, ser. UIST, Victoria, BC, Canada: ACM, 2009, pp. 121–124, ISBN: 978-1-60558-745-5.

[13] A. Guo and T. Paek, "Exploring tilt for no-touch, wrist-only interactions on smartwatches," in *Proceedings*, ser. MobileHCI, Florence, Italy: ACM, 2016, pp. 17–28, ISBN: 978-1-4503-4408-1.

[14] J. Gong, X.-D. Yang, and P. Irani, "Wristwhirl: One-handed continuous smartwatch input using wrist gestures," in *Proceedings*, ser. UIST, Tokyo, Japan: ACM, 2016, pp. 861–872, ISBN: 978-1-4503-4189-9.

[15] K.-Y. Chen, K. Lyons, S. White, and S. Patel, "Utrack: 3d input using two magnetic sensors," in *Proceedings*, ser. UIST, St. Andrews, Scotland, United Kingdom: ACM, 2013, pp. 237–244, ISBN: 978-1-4503-2268-3.

[16] K. Katsuragawa, K. Pietroszek, J. R. Wallace, and E. Lank, "Watchpoint: Freehand pointing with a smartwatch in a ubiquitous display environment," in *Proceedings*, ser. AVI, Bari, Italy: ACM, 2016, pp. 128–135, ISBN: 978-1-4503-4131-8.

[17] C. Zhang, Q. Xue, A. Waghmare, S. Jain, Y. Pu, S. Hersek, K. Lyons, K. A. Cunefare, O. T. Inan, and G. D. Abowd, "Soundtrak: Continuous 3d tracking of a finger using active acoustics," *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, vol. 1, no. 2, 30:1–30:25, Jun. 2017.

[18] A. Esteves, E. Velloso, A. Bulling, and H. Gellersen, "Orbits: Gaze interaction for smart watches using smooth pursuit eye movements," in *Proceedings*, ser. UIST, Charlotte, NC, USA: ACM, 2015, pp. 457–466.

[19] G. Reyes, J. Wu, N. Juneja, M. Goldshtein, W. K. Edwards, G. D. Abowd, and T. Starner, "Synchrowatch: One-handed synchronous smartwatch gestures using correlation and magnetic sensing," *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, vol. 1, no. 4, 158:1–158:26, Jan. 2018.

[20] M. Carter, E. Velloso, J. Downs, A. Sellen, K. O'Hara, and F. Vetere, "Pathsync: Multi-user gestural interaction with touchless rhythmic path mimicry," in *Proceed-*

ings, ser. CHI, San Jose, California, USA: ACM, 2016, pp. 3415–3427, ISBN: 978-1-4503-3362-7.

[21] J. Wu, C. Colglazier, A. Ravishankar, Y. Duan, Y. Wang, T. Ploetz, and T. Starner, "Seesaw: Rapid one-handed synchronous gesture interface for smartwatches," in *Proceedings*, ser. ISWC, Singapore, Singapore: ACM, 2018, pp. 17–20, ISBN: 978-1-4503-5967-2.

[22] J. Scott, D. Dearman, K. Yatani, and K. N. Truong, "Sensing foot gestures from the pocket," in *Proceedings of the 23nd annual ACM symposium on User interface software and technology*, ACM, 2010, pp. 199–208.

[23] J. Rico and S. Brewster, "Usable gestures for mobile interfaces: Evaluating social acceptability," in *Proc. CHI '10*, ACM, pp. 887–896, ISBN: 978-1-60558-929-9.

[24] H. Pohl, A. Muresan, and K. Hornbæk, "Charting subtle interaction in the hci literature," 2019.